

Article for Special Edition of the journal *Evaluation*

Editors of Special Edition: Justin Jagosh and Nick Tilley.

Realism and resources: Towards more explanatory economic evaluation

Rob Anderson, Associate Professor of Health Economics and Evaluation ^{a b}

Rebecca Hardwick, Associate Research Fellow and PhD candidate ^{a b}

^aInstitute of Health Research, University of Exeter Medical School, Exeter, UK

^bNIHR CLAHRC South West Peninsula, Exeter, UK

Corresponding author:

Prof. Rob Anderson

r.anderson@exeter.ac.uk

ESMI

South Cloisters

University of Exeter

St Luke's Campus

Exeter

EX1 2LU

United Kingdom

Tel: +44 1392 726085

Realism and resources: Towards more explanatory economic evaluation

Abstract

To be successfully and sustainably adopted, policy-makers, service managers and practitioners want public programmes to be affordable and cost-effective, as well as effective. While the realist evaluation question is often summarised as *what works* for whom, under what circumstances, we believe the approach can be as salient to answering questions about resource use, costs and cost-effectiveness - the traditional domain of economic evaluation methods.

This paper first describes the key similarities and differences between economic evaluation and realist evaluation. It summarises what health economists see as the challenges of evaluating complex interventions, and their suggested solutions. We then use examples of programme theory from a recent realist review of shared care for chronic conditions to illustrate two ways in which realist evaluations might better capture the resource requirements and resource consequences of programmes, and thereby produce explanations of how they are linked to outcomes (i.e. explanations of cost-effectiveness).

Word count: 142

Key words: realist evaluation, economic evaluation, cost-effectiveness, resource use, shared care, complex interventions

Introduction

The 'core hypothesis [of every programme] is always as follows: if we provide these people with these resources it may change these behaviours.'

(Pawson, 2003, p.472). *Nothing as practical as good theory.*

'There is likely to be a greater role for qualitative research methods in helping address many of the limitations of experimental approaches [including] to explore variation in users and interventions, and to better understand the nature of the interventions ... Furthermore, greater attention to how, why and for whom interventions work can help strengthen the validity of a study, by specifying the links between inputs and outcomes, and improve generalisability, through a better understanding of the context in which a particular intervention is likely to be cost-effective.'

(Byford and Sefton, 2003, p.107). *Economic evaluation of complex health and social care interventions.*

Like other theory-driven evaluation methodologies, the starting point for realist evaluation is to identify and express *how* and *why* the outcomes of social interventions or programmes are caused (Pawson, 2013, Pawson and Tilley, 1997, Pawson and Tilley, 2004). That is, such approaches tend to be based on expectations and explanations of greater effectiveness than alternative programmes or current service arrangements. Yet, to be successfully adopted and implemented, policy-makers, service managers and practitioners also want interventions and programmes to be affordable and cost-effective (in their locality, with their organisational structure, their professional staff mix, and within their budget etc.). The established evaluation methods which assess cost-effectiveness, or the efficiency of interventions, are economic evaluations (Drummond et al., 2005).

The primary question that drives realist evaluations is typically expressed as *what works for whom, under what circumstances and why*, and we believe the approach is as salient to answering questions about costs and cost-effectiveness (Anderson, 2004, Anderson and Shemilt, 2010). Realist evaluation makes prominent the role of invisible or intangible mechanisms, and we propose that the central conception of mechanisms as being both reasoning *and resources* implies that a potential focus on 'economic' mechanisms, contexts and outcomes – or on economic aspects of each of these - is consistent with the overall approach.

This paper aims to:

- Provide a brief recap of realist evaluation and a brief introduction to economic evaluation methods

- Describe the key similarities and differences between economic evaluation and realist evaluation.
- Provide an argument that most economic evaluations – at least as they are conducted in health care – are archetypal ‘black box’ evaluations, with minimal interest in how and why a particular configuration of resources (an intervention) changes outcomes, which has important consequences for the generalisability and use of their findings.
- Summarise what health economists see as the challenges of evaluating complex interventions, and therefore what approaches or adaptations should be adopted for conducting economic evaluations of them.

The paper will conclude by proposing the likely main features and practical challenges of realist economic evaluation and arguing that integrating explanations of resource use and cost-effectiveness can be done within realist evaluations.

While our research experience (and our worked example of shared care) draw on our backgrounds as health services researchers and economic evaluators, we hope that the arguments and illustrations made in this paper are accessible to evaluators in other sectors. Similarly, we hope that the paper will be interesting and accessible to both health economists curious about the implications of realist evaluation for enhancing the generalisability of their economic evaluation findings, and realist and other evaluators interested in extending their evaluations to explain costs and cost-effectiveness.

What are realist evaluations?

Realist evaluations are a form of theory-driven evaluation that aim to explain how policies, programmes and interventions work, who they work for, in what circumstances and why (Pawson, 2013, Pawson and Tilley, 1997, Pawson and Tilley, 2004). Recognising that no one policy, programme or intervention will always work, all the time, for everyone, realist approaches seek to explain this pattern of outcomes through building programme theories about how an intervention (policy or programme) is meant to work (often according to programme architects, or policy-makers, or participants), and then ‘test’ whether and how this programme theory plays out in the real world using empirical data. .

Within realist evaluation policies, programmes and interventions are not just a ‘treatment’, they are *theories incarnate*. The theories are the assumptions, perspectives, hunches and hopes in the policy maker’s head and often this theory is not articulated but remains hidden in the ‘black box’ of the intervention. Black box evaluations make no attempt to uncover or elucidate the causal connections between the inputs and tangible components of an

intervention and its expected outcomes (Lipsey, 1993). Theory-driven evaluation in contrast, including realist approaches, seek to 'open up the black box'; in order to "identify the causal processes that theoretically intervene between programme treatment and outcome" (Astbury and Leeuw, 2010). In this way, realist evaluation moves from describing what works, to explaining who it works for, in what context and why. Realist evaluation becomes an act of theory development and refinement.

Realist evaluation is based on a generative model of causation; that is, that there are mechanisms which bring about change, that exist independently of us being able to see them, but when circumstances are right, we see their effects, and from those observations can deduce what might be causing the change. In social policies, programmes and interventions, there is always an underlying reason why it is felt that a particular approach will be effective. What realist approaches aim to do is make explicit the underlying theory behind the intervention, and what the mechanism(s) is (or are) that is bringing about the change.

The methods used in realist evaluation are plural, as the focus is on developing and refining programme theory. Often quantitative data (e.g. from randomised controlled trials or surveys) are useful for describing the pattern of outcomes, and then qualitative data useful for understanding in more detail why something did or did not produce the expected outcomes.

What are economic evaluations?

In health care and other public service sectors economic evaluation is an umbrella term which encompasses a range of different evaluation methods, including cost-benefit analysis and cost-effectiveness analysis (Drummond et al., 2005, Layard and Glaister, 1994). As the hyphenated labels imply, all these evaluation methods involve the simultaneous comparison of both the costs and the value/impacts of the programmes or interventions being evaluated.

Another defining feature is that economic evaluations should involve the comparison of two or more programmes or interventions. They are fundamentally a decision-informing method or option appraisal tool, usually for a particular jurisdiction or policy maker at a given time. While in lay language it is possible to say something "is cost-effective", to economists and economic evaluators an intervention or programme can only ever be cost-effective relative to some other intervention or programme, or doing nothing (although, doing or changing nothing usually still has cost and other consequences).

How are they different from realist evaluation?

The approach of realist evaluation and the methods of economic evaluation could not be more different. Economic evaluation has evolved as a highly pragmatic decision-informing form of quantitative summative evaluation. While it is mostly conducted by people with a training in economics, it only partially draws upon a few key economic concepts and principles. The basis of economic evaluation methods in mainstream economic theory, in particular in welfare economics, are often claimed but also contested (Drummond and Maguire, 2001, Garber, 1996). Welfare economics is the branch of economic analysis which has developed principles and methods for informing societal (e.g. government) decisions that aim to maximise welfare at a societal level. For a good introductory description of welfare economics, and of the ways in which health economic evaluation methods are based on it, please see chapter by Tsuchiya and Williams, in Drummond and Maguire, 2001. However, for a contrasting view, that economic evaluation actually developed more pragmatically as a technique of applied engineering and decision analysis, Garber et al (1996) have said that “only recently have economists sought to graft cost-effectiveness analysis to theoretical roots in welfare economics.”(p.26).

We have compared the main features of realist evaluation and economic evaluation in Table 1 below. Looking at these many differences, it is unsurprising that any combined, or even complementary use, of the two approaches has only rarely occurred.

Table 1. Economic and realist evaluation compared

	Economic evaluation	Realist evaluation
Theoretical basis	Ostensibly, welfare economics; in practice more pragmatic and makes selective use of core economic concepts (e.g. cost-benefit principle, opportunity cost, marginal analysis)	Realism ^a : reality is both observable and external, and how it is interpreted
Conception of causality	Not generally known or discussed (but, implicitly, successionist causation: i.e. A causes B; seeking empirical regularities that intervention A is or is not cost-effective)	Generative notion of causation: outcomes are contingent on a range of underlying mechanisms, occur at different levels of reality, and will only ever only manifest as ‘demi-regularities’
Research aim	To produce estimates both of costs and effectiveness in a specific context (e.g. alongside a specific effectiveness study) or a particular decision-making jurisdiction (model-based economic evaluation)	To develop and refine programme theories (i.e. potential explanations) about how and why interventions work (i.e. produce beneficial outcomes), including how and why they work differently in different contexts

Policy aim	To inform specific decisions amongst a defined number of alternatives	To inform decision-makers about the way that the intervention produces its effects, and what modifies or influences that effectiveness
Type of data	Quantitative (especially: resource use, unit costs, final (e.g. health) outcomes). Secondary research data only as part of decision modelling	Quantitative and qualitative depending on the specific knowledge gaps. Often quantitative to establish the outcome pattern, and qualitative to determine how and why this pattern occurs. Secondary research (e.g. realist reviews) may complement primary research (e.g. for theory development stage)
Preferred study design?	Either: Experimental evaluation (e.g. randomised controlled trial) Or: Decision analytic (simulation) modelling, synthesising assumptions and cost and effectiveness evidence from variety of sources	Pluralist – no strongly preferred method or design (research question dependent) (although for some realist researchers, randomised experimental methods are seen as incompatible with realist conception of causation and mode of explanation)
Generalisability/transferability?	Emphasis on transferability of results (e.g. similar cost-effectiveness) contingent upon key features of context (country, patient group)	Generalisation is through progressively applying the programme theory to other contexts. Realist approaches recognise that the explanations developed from a realist evaluation are always open to further development and refinement.

Footnotes:

^a Originally, in Pawson and Tilley, 1997 (Chapter 3), the basis of realist evaluation was described as scientific realism. In the preface of his 2013 book Ray Pawson explains how his terminology has evolved to be 'some type of realism', while others may see realist evaluation as more closely aligned to, for example, critical realism.

Measurement vs explanation/theory-building

Perhaps the most notable difference between economic evaluation and realist approaches to evaluation is that realist approaches aim to build generalisable causal explanations of particular programmes, policies or services whereas economic evaluations essentially aim to measure or estimate the cost-effectiveness of introducing a particular programme. In the context of a trial-based economic evaluation this difference is most marked. The 'economic evaluation component' of a randomised trial will tend to only (a) collect data on the types and amounts of resource use (e.g. staff time, equipment, medicines, additional training) involved in providing the different interventions, and (b) collect data on any altered resource use associated with changed outcomes (e.g. patients with improved health should need less medication and fewer hospital or primary care visits). This is why some health economists

have criticised current methods of economic evaluation as unhelpfully intervention-focussed and outcomes-driven, often to the exclusion of considerations of context (Birch, 2002, Birch and Gafni, 2003).

Creating localised vs generalisable knowledge

There is a key difference between the primarily decision-informing goals of economic evaluation and the primarily theory- and generalisable knowledge-generating purpose of realist evaluation (Anderson and Shemilt, 2010). Model-based economic evaluations are always based on informing a well-specified policy choice in a given locality, country or other jurisdiction. Similarly, trial-based economic evaluations only yield a cost-effectiveness estimate in relation to the types of patients, interventions and settings included in that study (Sculpher et al., 2006). The generalisability of the findings of economic evaluations is known to be highly constrained by a wide range of contextual and other factors, and not least because resources cost different amounts in different places and times, and because the opportunity cost of particular resources will vary according to the decision context (and therefore the alternative potential use of those resources) (Anderson, 2010, Sculpher et al., 2004).

In summary, economic evaluations, as currently designed and conducted, are largely atheoretical exercises in measuring or estimating the inputs (resources and costs) and outcomes (effectiveness) of interventions and their comparators (Anderson, 2010, Birch, 2002, Lessard, 2007, Sculpher et al., 2004). While their full description of interventions and comparators is improving, they are often - in the health field at least - experimental and dominantly positivistic in approach; they can easily be characterised as 'black box' evaluations. That is, very few have any definite interest – or therefore any planned data collection - in the intervening causal mechanisms or relevant contexts that are thought to produce the expected improvements in outcomes or generate changes in resource use, *or link them to each other*. Thus, where there are several economic evaluations of 'the same' complex intervention or prevention programme, there is usually neither a clear result that one programme is consistently the most cost-effective in all contexts, nor evidence that can explain how and why the cost-effectiveness of intervention X or programme Y varies so much from context to context. This is what Pawson and Tilley referred to as 'the Martinson problem' – that nothing works (or is cost-effective) consistently - writ large (Pawson and Tilley, 1997) (p.30). A number of respected and experienced health economists share these views about the explanatory weaknesses of conventional economic evaluation methods, and also point to the need for more theory-driven approaches as the likely solution (Birch, 2002, Birch and Gafni, 2003, Coast, 2000, Drummond, 2010).

However, some economic evaluations are based on quite elaborate simulation models of long-term costs and outcomes of alternatives (decision models, or decision analytic models). While decision model-based economic evaluations are an apparent exception to the criticism that economic evaluations are black-box and atheoretical (because such simulation modelling at least provides an explicit framework of some of the supposed causal pathways and key trade-offs involved in a particular policy choice), they are not generally recognised as a form of theory development or theory-testing. This is in part because relations within model-based economic evaluations are mathematical expressions of assumed relationships, rather than causal propositions that could be discretely tested. Until the development of all decision models is based on more comprehensive and distinct processes of theory development, and perhaps can accommodate causal connections in non-mathematical ways and better capture the influence of contexts, then it seems to us they will only ever be a partial reflection of the theory-driven approach of realist evaluation.

Unsurprisingly, those health economists who have begun to confront the explanatory limitations of current methods of cost-effectiveness analysis, are those who have experience of conducting economic evaluations of complex interventions (such as models of service delivery, health policies or public health programmes). Below, in Box 1, we summarise the various challenges identified by health economists in evaluating complex health interventions.

Box 1. Challenges in the economic evaluation of complex health interventions

Conceptual challenges:

- Inevitability of multiple outcomes, including greater importance of some intermediate outcomes^a or non-health benefits^c Some important outcomes may not be expected or known at the beginning of an evaluation or may occur at system rather than individual level^e
- Difficulty attributing outcomes to interventions and/or transferring findings to other settings or populations with confidence^d
- Greater difficulty in defining appropriate alternatives for comparison (In particular, the issue of whether new services should be evaluated as alternatives to or additional to existing service arrangements)^a
- Greater importance of issues surrounding the scale of services (For example, in order for some types of community-based services to “become” cost-effective, they may have to handle enough patients for the alternative hospital service to be downsized)^b

Practical challenges:

- The use of established (e.g. clinical) outcome measures often not feasible^{a d}
- Incomplete identification of relevant costs, e.g. omission of capital costs^a
- Timing of evaluation of (newly introduced) services or programmes.^b Evaluation too early, and service operating under full capacity may underestimate effectiveness. Conversely, new or pilot projects may have higher levels of funding or staff enthusiasm/champions (e.g. investing extra time freely).^b Interventions in complex

systems may only create step-changes in outcomes after long periods^eThe cost of whole services or public health programmes often falls on different budgets. Cost savings in one area cannot always easily be re-invested elsewhere^d

Key to sources: a (Godber et al., 1997) b (Coast, 2000) c (Payne et al., 2013) d (Weatherly et al., 2009) e (Shiell et al., 2008)

In summary, most health economists have seen the challenges of evaluating complex interventions as ones which can be mainly tackled by either more comprehensive data collection or more sophisticated decision modelling or statistical techniques (Payne et al., 2013, Payne and Thompson, 2015, Sculpher, 2001). A few others, however, have acknowledged that the lack of conclusiveness and poor usefulness of economic evaluations is more deeply rooted in a dominantly positivist or clinical-epidemiological conception of how interventions generate outcomes, and which poorly recognises context, complexity or the active role of programme participants in producing outcomes (Birch, 2002, Byford and Sefton, 2003, Coast, 2000, Shiell et al., 2008). When health economists have suggested that greater use of theory might be part of the solution, they have typically advocated using complex systems theory in a broad sense rather than expanding the use of programme theory of the specific intervention type being evaluated (Lessard, 2007, Shiell et al., 2008).

What would more explanatory economic evaluation look like?

The following proposal for a more theory-driven approach to economic evaluation is based on discussions and presentations about this idea over the past five years, and also experiences of conducting a realist review of a complex intervention – shared care for chronic conditions, conducted by the present authors. In some ways it extends to empirical evaluation some of the realist-inspired ideas about conducting reviews of economic evidence suggested earlier. Our experiences suggest two main extensions to existing guidance on the design and conduct of realist evaluations.

First, we propose that the conception of mechanisms within realist evaluation, as a combination of both the *changed resources* offered and *participant's reasoning/responses* to those resources, be more fully and consistently articulated. This is because all mechanisms are to some extent economic (in the sense that they entail altered resources) but not all of these changes in resources have salient or substantial economic consequences (i.e. not all will involve opportunity costs, not all will require monetary investment). This first extension to the approach of realist evaluation is an endorsement of Dalkin et al's recent call to more fully disaggregate resources and reasoning when uncovering programme mechanisms (Dalkin et al., 2015), and that a realist evaluation that does this should go some way towards yielding explanations of cost and cost-effectiveness as well as effectiveness.

Second, we propose that there is value in identifying underlying ‘cost-effectiveness’-specific programme theories where they can be identified. By this we mean programme theories which explicitly seek to explain cost-effectiveness or altered costs. This will often be the case when, for example, a new model of care delivery (like shared care) is not necessarily believed to be *more effective* than current care, but rather is advocated because it is expected to be less costly while *not affecting effectiveness or safety etc.* In these situations, programme or service change designers will often express a number of ways in which they expect the new programme or service to be less costly or more efficient. These can often be captured as more generic programme theories, which we illustrate with the examples in the following section.

Programme theories of cost-effectiveness – a worked example

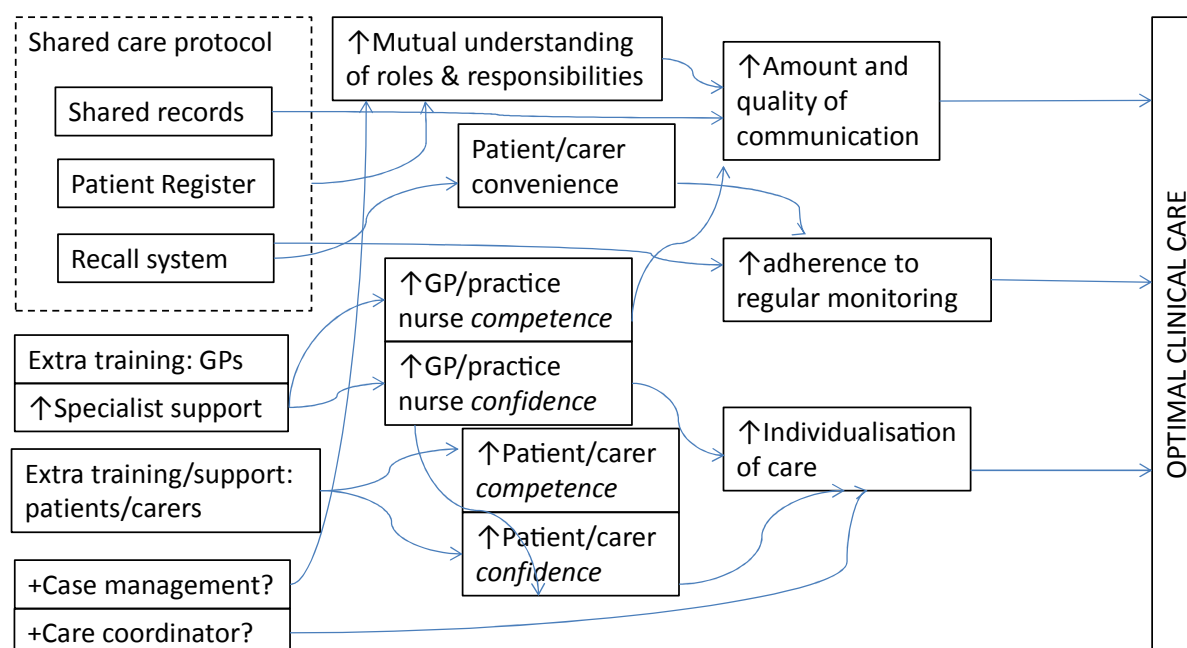
To illustrate the feasibility and value of developing cost-effectiveness-specific programme theories, we present selected findings from a recent realist review of the cost and cost-effectiveness of shared care we have conducted (Hardwick et al., 2013). While this was not a prospective realist evaluation of shared care, it demonstrates our main proposal that (a) discrete programme theories of cost-effectiveness can be identified in relation to a particular type of programme or health service (b) that these can sometimes be expressed in terms of mechanisms-contexts-outcome configurations, and (c) that it is possible to identify what outcomes, contexts or changes would need to be assessed in order to prospectively evaluate how and why a new shared care programme is or is not cost-effective.

Shared care is a way to manage the care of people in relation to a particular health condition, when the level of severity and/or stability of their condition means that optimal care requires the planned involvement of both primary care and specialist care professionals and services (Hickman M et al., 1994, Smith et al., 2007). It has been most used for patients with long-term chronic conditions such as asthma, diabetes, arthritis, high blood pressure and depression. A shared care programme will typically comprise the following components:

1. An agreed shared care protocol for the defined group of patients (which defines: the roles and responsibilities of primary care (e.g. GP) and disease specialist care professionals, and of patients and carers; the frequency and nature of monitoring or follow-up; and the rules or criteria for referral to different services or professionals)
2. Enhanced information exchange between disease specialists and generalists/primary care health professionals (computerised, paper or telephone)
3. Care coordination (computer systems and/or dedicated person)
4. Extra education and/or training (e.g. for GPs in the specific chronic disease, or patients in self-care)

This can be summarised graphically, as a flow diagram - including some speculative causal pathways between the core components of shared care and potential intervening 'mechanisms' which link the components to expectations of optimal care for a given patient group (Figure 4).

Figure 1. Provisional 'causal map' of shared care components, mechanisms and outcomes



To identify programme theories of the effectiveness and cost-effectiveness of shared care we conducted a review of several seminal descriptions of shared care and other descriptive and evaluation studies of actual examples of shared care. Typically, claims and assertions about how and why shared care was expected to be more effective or more cost-effective than usual care were in the introduction or discussion sections of empirical evaluation studies. Two main programme theories of the effectiveness of shared care were identified: the systematisation of high quality *care processes*, and; the fostering of positive and reciprocal *working relationships*. A three-part programme theory relating to expected efficiency gains due to introducing shared care (i.e. cost-effectiveness) was expressed as follows:

For defined types of patient, if care is shared in a systematic way between particular organisations or clinicians working in primary (generalist) and secondary (disease specialist) care, then there are efficiency gains within the health system due to:

- (a) Shifting care to less costly settings and/or less expensive practitioners

- (b) Better tailoring of resources to need
- (c) Enhancing use of patient and carer resources (and thereby substituting for formally provided care services)

These programme theories of the cost-effectiveness of shared care could therefore be labelled as ‘shifting’ (to less expensive care settings), ‘tailoring’ (to lowest needed or optimally cost-effective level of care) and ‘substitution’ (to care which is effectively ‘free of charge’ to the health system).

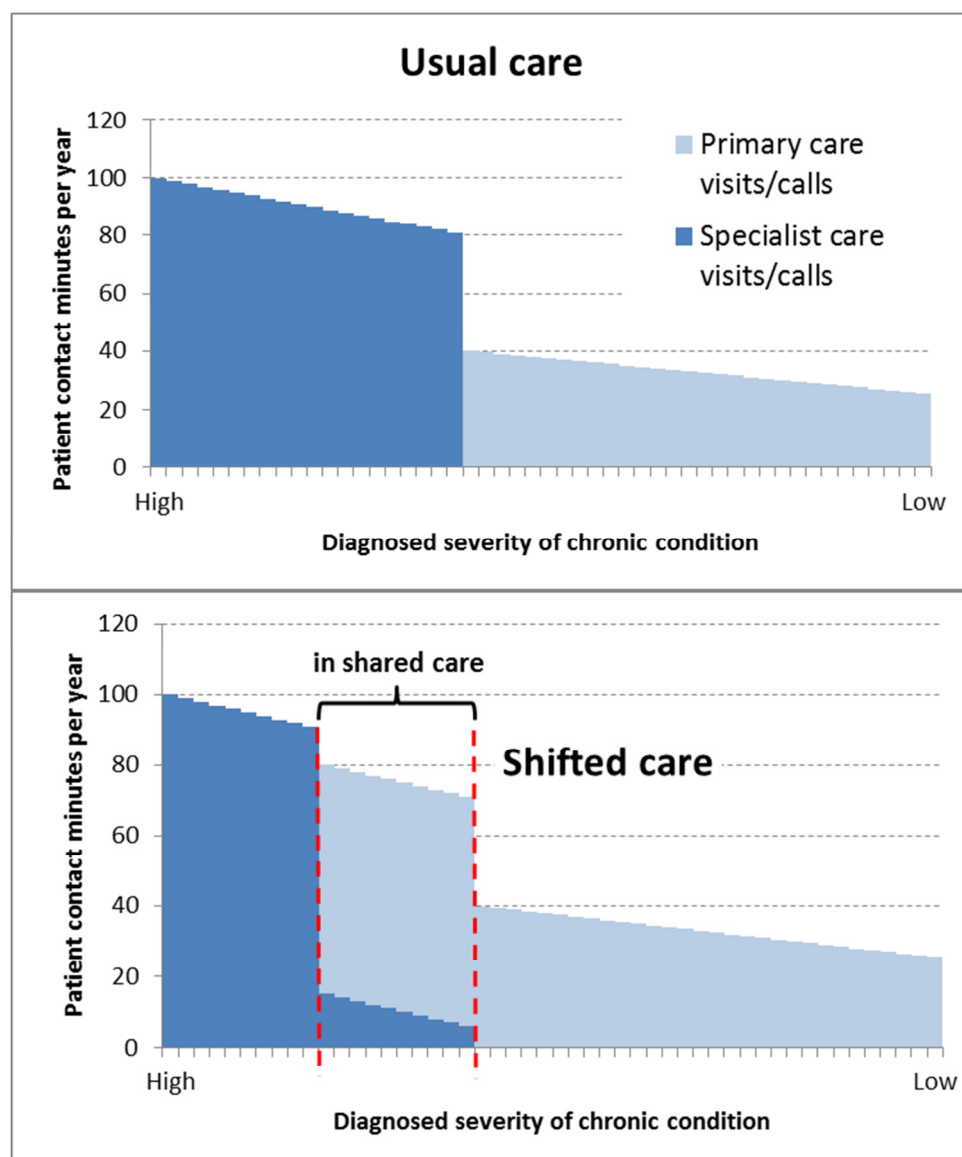
It should be noted here that, within the methods of economic evaluation, cost-effectiveness is a compound concept – so programme theories of cost-effectiveness should explain how *both* costs and effectiveness are produced *and* how they are causally linked. Some economic evaluations may dominantly focus on cost differences, but this is typically where equal or equivalent effectiveness has been assumed on the basis of existing evidence. In relation to particular models of care, the terms ‘cost-effective’ or ‘efficiency gains’ therefore would generally include: care which both costs less and improves health outcomes; care which costs less while maintaining the same outcomes; or care which improves outcomes at acceptable and affordable additional cost.ⁱ

These three programme theories of shared care can be expressed graphically, as shown in Figure 2, Figure 3 and Figure 4 below. Each graph shows total clinical contact minutes per year (i.e. staff resource consumption) against the severity of a patient’s chronic disease (running from high severity at graph origin, to lower severity – and where severity is really shorthand for any aspects of a patient’s condition that justifies more specialist input). The absolute values are not meant to be accurate, but merely to illustrate some interrelationships.

Under **usual care** (Figure 2), there is typically a positive association between disease severity and the amount of clinical contact time a patient receives – on average, the more severe or unstable a patient’s condition the more clinical contact time they should receive. Also, above a particular threshold level of severity it is assumed the patient’s clinical care for this chronic condition would shift from being wholly managed in primary care to being wholly managed in secondary/specialist care (e.g. regular hospital outpatient/ambulatory care visits to a consultant diabetologist, respiratory physician or psychiatrist). This diagram also assumes that being wholly under specialist clinical care would entail more clinical contact minutes during a typical year (perhaps because routine GP or practice nurse appointments are usually shorter than the typical hospital specialist outpatient appointment, and/or because more severe or unstable patients may require more frequent follow-up).

With the programme theory of **shifted** care, there is a group of patients with moderately severe disease –who under usual care would be cared for only by disease specialists - whose optimal care involves mainly primary care clinical visits but with the care protocols, enhanced information exchange and care coordination of shared care requires much less frequent contact with disease specialists (Figure 2). These less frequent contacts with specialists might be either planned (e.g. following initial referral/diagnosis and for initial treatment planning, or at regular review intervals), or unscheduled instances where symptoms change or treatment needs reviewing for other reasons (e.g. referred ‘back’ to the specialists according to criteria in the shared care protocol). In turn, if specialist appointments are provided by more highly paid clinicians (e.g. hospital consultants) or in more expensive settings (e.g. hospitals) than GPs/nurses working in primary care, then this shift in the combination of resources used would lead to cost savings (at least on the basis of changes in staff mix/ clinical care setting, if ignoring any additional costs of introducing and supporting shared care).

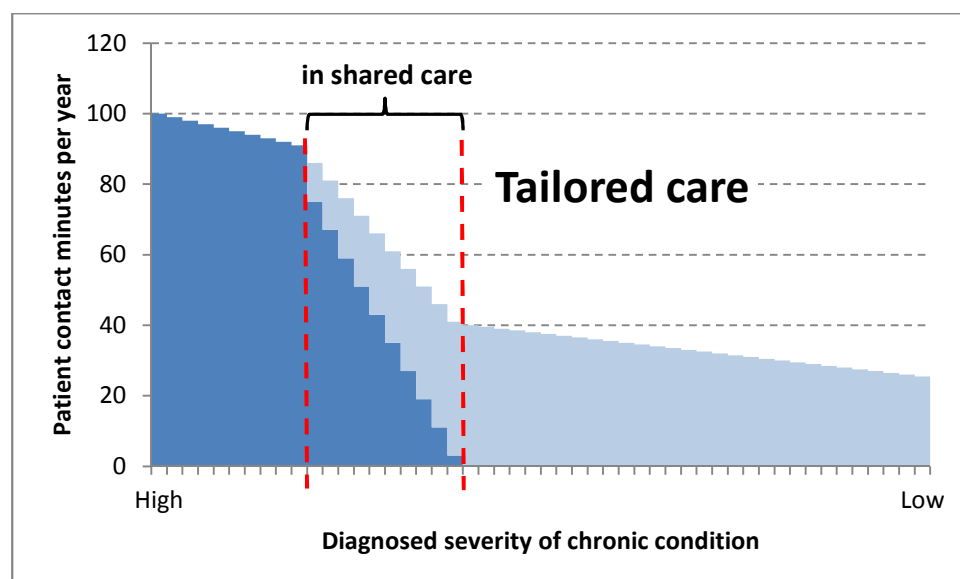
Figure 2. Usual care and ‘shifted’ care: hypothetical clinical contact time by disease severity



If shared care is more cost-effective through care being more ***tailored*** to need, then it reflects that for a group of patients with moderately severe disease the optimum mix of specialist care and primary care input and contact time can vary widely – anything between almost all clinical care being provided by disease specialists, to almost all care being provided by the GP or other primary care practitioners. In other words, there is a more gradual reduction of specialist input to care for patients who have less severe or more stable disease, while at the same time the level of care from primary care increases (until, perhaps, specialist care is only required when treatment fails or symptoms worsen considerably). Note that this graph only shows the staff time resource aspects of this possibility. In the

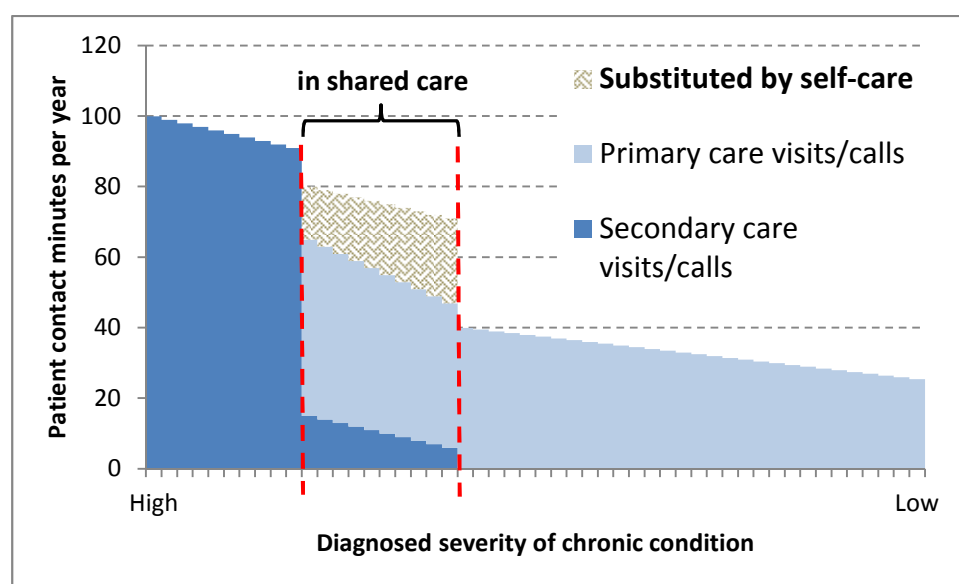
longer term, the cost-effectiveness of tailoring also depends on the degree to which greater input from disease specialists actually leads to more effective treatment monitoring and decisions in relating to changing or complex clinical needs.

Figure 3. ‘Tailored care’: clinical contact time by disease severity



Our third programme theory to explain the potential efficiency gains of shared care asserts that greater self-care and self-monitoring by patients would reduce the need for formal care and monitoring appointments from a health professional (i.e. **substitution**). If such resource substitution takes place, and the evaluation aims to assess costs to the health or public sector, then there would be directly reduced costs of care services. Figure 4 shows the potential savings due to increased self-care in addition to those of shifted care. It also shows such substitution as only affecting those patients under shared care; but of course there might be ‘spillover effects’, if some of the extra support and improved skills for self-care and home monitoring is given to all patients under the care of the same GPs or primary care practices.

Figure 4. ‘substituted care’: clinical contact time by disease severity



Refining the theories of cost-effectiveness

In this section we describe what evidence we discovered in published economic evaluations of shared care in relation to two of these three programme theories: care shifting and care tailoring. While this evidence comes from conventional economic evaluations which were not explicitly theory-driven or realist in approach, we believe these worked examples usefully illustrate how prospective studies could generate relevant data to refine such programme theories in a particular health systems setting and patient group, and with a particular specification of shared care.

Cost-effectiveness by care shifting - evidence and key contexts

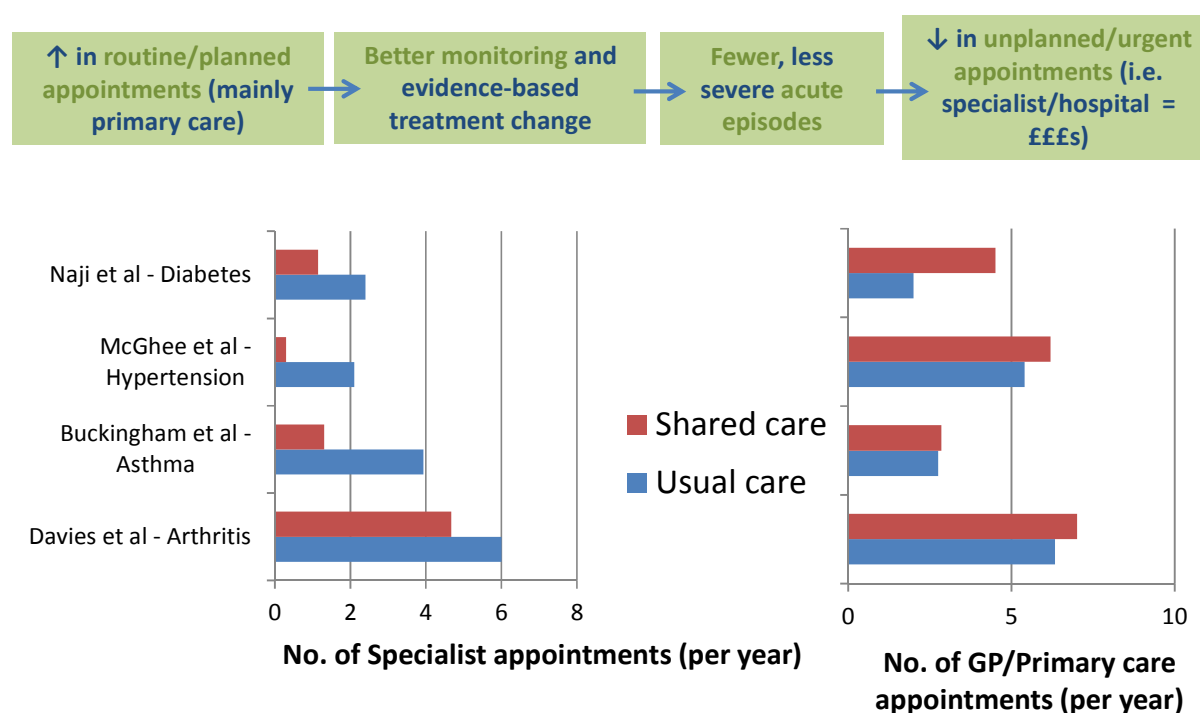
The shifting of care to less costly care professionals or to less costly clinical settings appears to occur in two main ways within shared care. Firstly, many shared care protocols stipulate a recommended frequency of routine follow-up tests and visits to monitor a patient's health and, where necessary, adjust treatment. This typically involved nearly all follow-up appointments in primary care, with a GP or practice nurse, rather than regular follow-up a hospital-based specialist. So, for example, in a cost-effectiveness analysis of shared care for people with hepatitis C in Queensland Australia, usual specialist-based follow-up required 15 hospital appointments over a period of a year (Anderson & Haas, 2001). Under shared care, there were still 15 follow-up visits scheduled, but nine of them would be with the patient's GP. If attendance at appointments is beneficial, and GP appointments are considerably cheaper than hospital specialist appointments (as is often the case), then such savings are effectively 'hard-wired' into the shared care protocol.

Note the possible intermediate mechanism of adherence to attending appointments, which may modify the cost differences due to shifting care, or increase effectiveness if greater follow-up appointment adherence leads to more effective monitoring and treatment.¹ In the evaluation of shared care in Queensland, interview evidence strongly suggested that follow-up by more geographically local (and possibly better known or trusted) GPs had led to substantially higher levels of follow-up and adherence to treatment. Such regular follow-up was especially critical given the potential psychiatric side-effects of the drugs being taken by the patients with hepatitis C.

The cost-effectiveness studies of shared care also suggested another extension of the underlying programme theory of shifting care, which involves a shift in the balance of planned versus unplanned care. As part of the planned shift to primary care appointments, other supporting aspects of shared care such as a patient register and recall system often increase planned attendance for monitoring and follow-up overall. This should, in theory, lead to better monitoring and any necessary treatment changes, and therefore a better managed condition. This in turn can lead to fewer or less severe acute episodes (e.g. asthma attacks, hypoglycaemic episodes, episodes of depression), and an associated reduction in unplanned or urgent appointments primary care or hospital appointments. The hypothetical causal chain of this more specific version of the care shifting programme theory is shown in Figure 5 below.

But does the evidence from economic evaluations support this theory? For the economic evaluations of shared care in the UK in the 1990s, this seemed to be the case. While there was not evidence relating to all of these steps in all studies (most did not distinguish planned from unplanned visits), where the comparator was usual hospital care then with shared care primary care visits increased while specialist appointments decreased (see bar charts in Figure 5).

Figure 5. Cost savings due to a better managed condition



These lower numbers of specialist or emergency care appointments, probably through better planned monitoring and treatment changes, might lead to considerable cost savings, especially if any urgent treatment for acute episodes also usually occurred in hospital or with specialists. But the differences between the cost of a specialist appointment in hospital and a GP appointment were not always large, so this might reduce the cost savings associated with these shifts in appointment patterns.

Sometimes however, shared care represented a shift towards *more*, rather than *less*, disease specialist involvement; in these cases shared care was meant to provide higher access to specialists for patients from primary care, and thereby greater costs. This was often the case with shared care for depression in the United States (often called 'collaborative care'), where the comparator in evaluations was typically usual primary physician care. Unsurprisingly, this was associated with higher in costs – albeit often with improved outcomes (between 30% higher and double the cost of usual care in such studies (Pyne et al., 2010, Rost et al., 2005, von Korff et al., 1998)). Thus, in the context of previously *under-met* need for disease specialist care, shared care was often cost-increasing.

Cost-effectiveness by tailoring care - evidence and key contexts

In two of the evaluations of shared care for mental health in the United States, better tailoring of care to needs was an explicit aim of shared care (Rost et al., 2005, von Korff et al., 1998). In Rost et al. this was expressed more as what could be called 'temporal tailoring'; more regular scheduled contact with patients was primarily to encourage 'continued treatment adherence when symptoms were resolving, to adjust treatment if symptoms were not resolving, and to terminate treatment which patients in remission did not require' (p8). These studies suggested this was especially important for effective treatment of people with depression.

In von Korff et al, the expression of tailoring was less explicit and (in the care model aims) more couched in the terms of 'improved management' (von Korff et al., 1998). However, there was also evidence of a care tailoring mechanism at work: 'Collaborative Care patients were more likely to be switched to new medications, often a selective serotonin reuptake inhibitor [SSRI], as their care progressed because of the closer follow-up that Collaborative Care patients received' (p.145). So, in this case (assuming the previous medications were older, cheaper and less effective than SSRIs) tailoring could be viewed as 'tailoring up'.

The potential patient benefits of tailoring could be limited or non-existent if the actual range of treatments or levels of treatment were not available. For example, in the Pyne et al., study, treatment tailoring was limited to altering medication, without any opportunities for offering psychological or counselling based therapies (for which there are known patient preferences). Also, in the same study, the authors believed that the effectiveness of stepped collaborative care model was limited because it did not tackle common co-morbidities of depression (such as pain, anxiety and substance abuse; p.819, (Pyne et al., 2010))

In some studies there was some data on the measures of variance (e.g. standard error/deviation) for numbers of visits or other units of service use. Wider variances might provide indirect support that shared care is leading to more tailoring – that is, perhaps under shared care, the same patient group show wider variation in their service usage rates (implying that either some are getting more or some are getting less than under usual alternative care). Other, more direct, evidence of tailoring would come from documenting the frequency of treatment changes, or changes in the pattern of follow-up monitoring. Whether driven by clinical considerations or by patient preferences, better tailoring to need seems to be a well-recognised mechanism of many models of health care improvement. However, better tailoring of care to need can only, by definition, be as cost-effective as the specific treatments and monitoring strategies that people get at each 'level' of care provided.

Conclusions and recommendations

Our experience from conducting our systematic review of the cost-effectiveness of shared care suggests two complementary approaches for using the logic of realist inquiry to explain differences in costs and cost-effectiveness between different models of care, or other complex interventions. First, we have shown that for some types of programme or service change, it is possible to specify cost-effectiveness-specific programme theories. In fact, some of these programme theories of cost-effectiveness – for example, better tailoring of care to need – are so closely linked to notions of efficiency that it could be argued they are more like Pawson's notion of transferrable 'conceptual platforms', and so they might be applicable to explaining the cost-effectiveness of a wide range of programmes or service changes (Pawson, 2013).

As well as being able to specify such cost-effectiveness-specific theories for shared care, we have shown that it is possible to express some of them in terms of mechanisms, relevant contexts and particular outcomes (i.e. the basic components of realist explanation-building). A possible exception to this might be when care (e.g. follow-up appointments) is directly shifted to less costly settings or practitioners, or when some component of an intervention (e.g. access to a specialised drug) is both very costly and unique to the new intervention. Arguably, to explain the resulting cost difference due to 'shifted care' or similar situations requires not so much theory as basic arithmetic. However, we believe such deterministic, linear 'hardwiring' of cost differences will not be a feature of all complex interventions, and many other determinants of resource use will be more multi-level, variable, behaviour-dependent and more context-sensitive and therefore require a more realist logic of inquiry to elucidate them. As well as our example, others have begun to use cost-effectiveness specific programme theories within realist reviews (Charles et al., 2013, Charles, 2014) or using realist evaluation to explain 'economic interventions', such as policies regarding fees to access care (Marchal et al., 2013).

The second approach, which could complement the development and refinement of cost-effectiveness-specific programme theories, involves raising recognition that all three of the main components of realist explanations – contexts, mechanisms and outcomes – can either *require* additional resources in order to exist, or *impact* on resources. This endorses and extends Dalkin et al's recent argument that the realist notion of mechanism should more clearly disaggregate the mechanism reasoning and mechanism resources. Where realist evaluators may need to learn more, is how resources consumed in different decision contexts, or at different overall volume, affects whether and how much there is actually an opportunity cost for those resources.

Considering these suggested approaches together, we do not suggest a new sub-genre of 'realist economic evaluation'. This would, we believe, unhelpfully and artificially separate evaluations and explanations of effectiveness from cost-effectiveness (including explanations of changing patterns of resource use). Our main conclusion is that these goals will always be inextricably linked and are therefore best evaluated together. In a way, the common realist expression of what a programme theory is (reasoning *and resources* – as captured by the Pawson quotation at the beginning of this paper) already firmly acknowledges this. Instead, we think that evaluations could more explicitly theorise and capture the resource requirements and consequences of hypothesised programme mechanisms, outcomes and contexts. Such process evaluations are now increasingly encouraged and funded alongside effectiveness evaluations of complex interventions, including greater endorsement that they should be theory-driven, at least in the UK (Moore et al., 2014)

Another conclusion from our review and analysis, is that the cost-effectiveness of generic models of care like shared care can only be as cost-effective as the underlying treatments and chronic disease monitoring strategies around which their protocols are built. If 'stepping up' a patient with depression to a stronger and more expensive drug and more frequent monitoring by psychiatric specialists does not reflect additional effectiveness of that new treatment, and better management of side-effects through the closer monitoring, then in this case the cost-effectiveness of sharing care will be undermined. Unfortunately, in different chronic disease areas even the knowledge about the relative cost-effectiveness of existing treatments is often poor.

For existing realist evaluators, this is neither rocket science, nor hardcore economics; for example, the hypothetical 'mechanism' of more experienced practitioners making quicker yet more effective treatment decisions, will have the expected resource consequence of shorter clinical encounters, but at the expense of higher pay rates per minute as the additional (or marginal) 'cost' of applying this greater clinical experience. These things can often be measured, or at least observed at some level. However, other aspects of complex health and social interventions do raise additional quite specific challenges for understanding resource use and economic evaluation. Issues of scale affect both effectiveness and costs. Varying decision contexts (including different budget constraints) also mean that the same services or programmes can have quite different opportunity costs in different places and populations. Furthermore, people in different countries and cultures may value health, non-health and equity outcomes differently (especially where the primary outcome may be information, rather than health or other more measurable outcomes).

Progress in this area of evaluation more fundamentally needs a larger community of economic evaluators and economists willing to learn about and try out realist approaches in their evaluative studies. Equally, progress also needs more realist evaluators who are both undaunted by the jargon of economic evaluation methods and keen to fully theorise the role resources play in intervention mechanisms, contexts and outcomes. While our insights and this paper have drawn upon our own experiences of trying to understand complex health interventions, the imperative of cost-effectiveness and efficiency as goals is also central to other areas of policy evaluation and organisational management. In these other policy areas, the methodological habits, home disciplines and epistemological starting points of economic evaluators may make them more amenable to using realist and theory-driven evaluation approaches than seems to be the case in health and medical research (where, arguably, the experimental imperatives of clinical epidemiology have been the dominant methodological influence on economic evaluation methods). And of course, as well as extending knowledge and skills into perhaps unfamiliar territory, evaluators will also simply need more time or access to additional data sources in order to broaden their focus to resources and costs; only then will they be able to evidence the underlying mechanisms and contexts that seem to explain patterns in them. Research funders too will need to recognise this. This way we might increase confidence that full exploration of programme theories of effectiveness should produce explanations of cost differences and cost-effectiveness as an inevitable and valuable by-product.

References

- ANDERSON, R. 2004. Realistic Evaluation [book review]. *Applied Health Economics and Health Policy*, 2.
- ANDERSON, R. 2010. Systematic review of economic evaluations: utility or futility ? *Health Economics*, 19.
- ANDERSON, R. & SHEMILT, I. 2010. The role of economic perspectives and evidence in systematic review. In: SHEMILT, I., MUGFORD, M., VALE, L., MARSH, K. & DONALDSON, C. (eds.) *Evidence-based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice*. 2nd ed. Oxford, UK: Wiley-Blackwell.
- ASTBURY, B. & LEEUW, F. 2010. Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31, 363-381.
- BIRCH, S. 2002. Making the problem fit the solution: evidence-based decision-making and 'Dolly' economics. In: DONALDSON, C., MUGFORD, M. & VALE, L. (eds.) *Evidence-based health economics: from effectiveness to efficiency in systematic review*. London: BMJ Books.
- BIRCH, S. & GAFNI 2003. Economics and the evaluation of health care programmes: generalisability of methods and implications for generalisability of results. *Health Policy*, 64, 207-219.
- BYFORD, S. & SEFTON, T. 2003. Economic Evaluation of complex health and social care interventions. *National Institute Economic Review*, 186, 98-103.
- CHARLES, J. 2014. Reflections on using realist methodology to explore and explain costs and cost-effectiveness - the FEMuR Feasibility Trial. *CARES Conference*. Liverpool.
- CHARLES, J., EDWARDS, R., WILLIAMS, N., DIN, N. & ON BEHALF OF THE FEMUR REALIST REVIEW STUDY GROUP 2013. The application of realist synthesis review methods in public health economics [Meeting abstract]. *Lancet*, 382, S28.
- COAST, J. H., M. MULLIGAN, J. SHEPPARD, S. JONES, J. 2000. Conceptual and practical difficulties with the economic evaluation of health services developments. *Journal of Health Services Research and Policy*, 5, 42-48.
- DALKIN, S., GREENHALGH, J., JONES, D., CUNNINGHAM, B. & LHUSSIÉ, M. 2015. What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science*, 10.

- DRUMMOND, M. 2010. Evidence-based decisions and economics: an agenda for research. *In: SHEMILT, I., MUGFORD, M., VALE, L., MARSH, K. & DONALDSON, C. (eds.) Evidence-Based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice*. 2nd ed. Oxford, UK: Wiley-Blackwell.
- DRUMMOND, M. & MAGUIRE, A. 2001. *Economic evaluation in healthcare: Merging theory with practice*, Oxford, Oxford University Press.
- DRUMMOND, M., SCULPHER, M., TORRANCE, G., O'BRIEN, B. & STODDART, G. 2005. *Methods for the economic evaluation of health care programmes*, Oxford, Oxford University Press.
- GARBER, A. 1996. Theoretical foundations of cost-effectiveness analysis. *In: GOLD, M., SIEGEL, J., RUSSELL, L. & WEINSTEIN, M. (eds.) Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.
- GODBER, E., ROBINSON, R. & STEINER, A. 1997. Economic evaluation and the shifting balance towards primary care: definitions, evidence and methodological issues. *Health Economics*, 6, 275-294.
- HARDWICK, R., PEARSON, M., BYNG, R. & ANDERSON, R. 2013. The effectiveness and cost-effectiveness of shared care: protocol for a realist review. *Systematic Reviews*, 2.
- HICKMAN M, DRUMMOND N & GRIMSHAW J 1994. A taxonomy of shared care for chronic disease. *Journal of Public Health Medicine*, 16, 7.
- LAYARD, R. & GLAISTER, S. 1994. *Cost-benefit analysis*, Cambridge, Cambridge University Press.
- LESSARD, C. 2007. Complexity and reflexivity: two important issues for economic evaluation in health care. *Soc Sci Med*, 64, 1754-65.
- LIPSEY, M. 1993. Theory as method: Small theories of treatments. *New Directions for Evaluation*, 1993, 5-38.
- MARCHAL, B., VAN DER VEKEN, K., ESSOLBI, A., DOSSOU, J.-P., RICHARD, F. & VAN BELLE, S. 2013. Methodological reflections on using realist evaluation in a study of fee exemption policies in West Africa and Morocco. FEMHealth Project, University of Aberdeen.
- MOORE, G., AUDREY, S., BARKER M, L, B., , B. C. & HARDEMAN W, M. L., O'CATHAIN A, TINATI T, WIGHT D, BAIRD J 2014. Process evaluation of complex interventions:

Medical Research Council guidance. London: MRC Population Health Science Research Network.

PAWSON, R. 2003. Nothing as practical as good theory. *Evaluation*, 9, 471-490.

PAWSON, R. 2013. *The Science of Evaluation: a realist manifesto*, London, Sage publications.

PAWSON, R. & TILLEY, N. 1997. *Realistic Evaluation*, London, Sage publications.

PAWSON, R. & TILLEY, N. 2004. Realist evaluation.

PAYNE, K., MCALLISTER, M. & DAVIES, L. 2013. Valuing the economic benefits of complex interventions: When maximising health is not sufficient. *Health Economics*, 22, 258-271.

PAYNE, K. & THOMPSON, A. 2015. Economic evaluations of complex interventions. In: RICHARDS, D. & RAHM HALLBERG, I. (eds.) *Complex interventions in health: An overview of research methods*. Abingdon, Oxfordshire: Routledge.

PYNE, J., FORTNEY, J., TRIPATHI, S., MACIEJEWSKI, M., EDLUND, M. & WILLIAMS, D. 2010. Cost-effectiveness analysis of a rural telemedicine collaborative care intervention for depression. *Archives of General Psychiatry*, 67, 812-821.

ROST, K., PYNE, J., DICKINSON, L. & LOSASSO, A. 2005. Cost-effectiveness of enhancing primary care depression management on an ongoing basis. *Annals of Family Medicine*, 3, 7-14.

SCULPHER, M. 2001. Economic evaluation. In: FULOP, N., ALLEN, P., CLARKE, A. & BLACK, N. (eds.) *Studying the organisation and delivery of health services: research methods*.

SCULPHER, M., CLAXTON, K., DRUMMOND, M. & MCCABE, C. 2006. Whither trial-based economic evaluation for health care decision making? *Health Economics*, 15, 677-687.

SCULPHER, M., PANG, F., MANCA, A., DRUMMOND, M., GOLDER, S., URDAHL, H., DAVIES, L. & EASTWOOD, A. 2004. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technology Assessment*, 8, 1-192.

SHIELL, A., HAWES, P. & GOLD, L. 2008. Complex interventions or complex systems? Implications for health economic evaluation. *British Medical Journal*, 336, 1281-1283.

SMITH, S. M., ALLWRIGHT, S. & O'DOWD, T. 2007. Effectiveness of shared care across the interface between primary and specialty care in chronic disease management. *Cochrane Database Syst Rev*, CD004910.

- VON KORFF, M., KATON, W., BUSH, T., LIN, E., SIMON, G., SAUNDERS, K., LUDMAN, E., WALKER, E. & UNUTZER, J. 1998. Treatment costs, cost offset, and cost-effectiveness of collaborative management for depression. *Psychosomatic Medicine*, 60, 143-149.
- WEATHERLY, H., DRUMMOND, M., CLAXTON, K., COOKSON, R., FERGUSON, B., GODFREY, C., RICE, N., SCULPHER, M. & SOWDEN, A. 2009. Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. *Health Policy*, 93, 85-92.

Footnote

ⁱ At a health system level, efficiency gains might also be seen to arise if *lower* health outcomes are achieved alongside yielding sufficient resource and cost *savings*, which would – or at least could - be reinvested elsewhere in the system *to achieve greater health gains than those lost* – for example in a different service and patient group. However, in reality, for service reorganisation interventions in established publicly funded health systems, providing care of reduced effectiveness is generally politically unacceptable and more unlikely).